Income and Unemployment Effects on Life Insurance Lapse*

Adjmal S. Sirak†

October 2015

Abstract

Using a large representative household panel we estimate hazard rate models to analyze the lapse behavior in the German life insurance market. Being the first study to combine detailed socio-demographic information with survival analysis techniques, we re-examine established hypotheses on the reason to lapse. We find significant effects of household income, wealth proxies, and unemployment. In contrast to previous studies, policyholder's age and gender have no effect once the former variables are controlled for.

Keywords: Life insurance lapse, emergency fund hypothesis, survival analysis

JEL-Classifications: G11, G22

*We are grateful to Martin Eling, Torben Kuhlenkasper, Andrey Launov, Thomas Otter, J. François Outreville, and to seminar participants at Johann Wolfgang Goethe University Frankfurt, University of St. Gallen as well as Nanyang Technological University Singapore for helpful comments. Elia Berdin, Philipp Krohn, Dominik Mattar, Tobias Niedrig, Stefan Perleth, and Manfred Poweleit assisted with their institutional knowledge. We are also grateful to Muhammed Altuntas, Jörg Finsinger, and a German market research institute for providing us with data. Timon Koehler provided invaluable research assistance. Financial support from ING-DiBa AG is gratefully acknowledged. Furthermore, we thank GfK SE for providing us with the data.

†Contact author. Johann Wolfgang Goethe University Frankfurt, Germany.
Email: adjmal.sirak@finance.uni-frankfurt.de

1. Introduction

In this paper we investigate the effects of variation in income and wealth on the propensity to lapse a life insurance policy. Life insurance lapse is relevant for insurance managers, because lapse rates affect a company's liquidity and profitability. It is relevant for regulators, as particularly mass lapse events may affect aggregate financial stability. Lapse risk therefore significantly affects insurers' solvency capital requirements under Solvency II. Furthermore, to lapse a life insurance may have adverse effects on the policyholder's wealth in case the surrender value is low relative to the value of a continued policy, as it is typically the case, particularly in the first years after inception of the contract.

We use a large and representative household panel to study how various socio-demographic variables, notably income and wealth, as well as changes in households' income and employment status affect policyholders' propensity to lapse. We can further estimate hazard rate and survival functions across policies that are representative for the entire population and span a large period of time. As further discussed below, in contrast to almost all of the extant literature this is made possible as our data combines the following information: 1) Households in the panel report income, the holding of other financial assets, as well as employment status; 2) the reported income comprises that of all household members and from all sources; 3) over the reporting period, from 2005 to 2011, new contract inceptions as well as cancellations are reported on a quarterly basis, while for all contracts held by households the firm name as well as date of inception are reported.

Endowment life insurance products, which is the object of our analysis, combine the features of a term life insurance and a regular savings vehicle. Policyholders make repeated contributions, usually at monthly frequency. Time to maturity may be chosen individually and typically ranges between twelve and 30 years. In case the policyholder dies before maturity, a predetermined amount is paid out to the insured party, e.g. family members. In case the policyholder survives through maturity, the invested capital including returns is paid out to her. The policyholder has the right to terminate the contract. In this case she is entitled to the surrender value, which may however result in a large loss relative to the scenario where the policy was retained until maturity. Specifically, whenever a policy contribution is paid, only part of it contributes to the invested capital, while the remainder is deducted to cover management fees and, especially in the first years, sales agents' commissions. This may result in the surrender value being zero or close to zero if the contract is lapsed shortly after inception, in which case the insurance contract as a whole becomes a negative return investment for the policyholder. But even later in its life, a policyholder who terminates the contract typically foregoes part of the expected return compared to policyholders who serve the contract till maturity.

We focus our analysis on the so-called emergency fund hypothesis (EFH). This asserts that in the face of financial distress policyholders either want to access the surrender value or free the funds that would otherwise go to policy contributions. Our analysis strongly supports this hypothesis. Notably, we find strong effects of income as well as for proxies for households' financial wealth. Moreover, our panel analysis reveals how changes in employment status trigger lapses. Our quantitative analysis may thus inform both regulators as well as insurance companies when conducting stress tests with respect to lapse behavior, notably following a sharp deterioration of macroeconomic conditions and a resulting surge in unemployment as well as a substantial drop in household income. Our main results are derived using the Cox proportional hazards model, though we also run a pooled panel logit regression, which confirms our results. Furthermore, we employ a loglogistic regression to quantify marginal effects on survival probabilities and to conduct a scenario analysis. In terms of magnitude, amongst other results we show that transition to unemployment raises the lapse rate[1] by more than 75%.

The literature has analyzed also other lapse motives. According to the policy replacement hypothesis (PRH) policyholders may lapse existing contracts to switch to a more attractive competitor. This opportunity cost argument is generalized by the interest rate hypothesis (IRH), which states that policyholders may decide to lapse their life insurance if market rates are sufficiently high. Our panel information on policyholders' subsequent decisions, following a lapse, allows us to safely rule out the replacement hypotheses for most of the reported lapses. Also, as we further discuss below, market circumstances as well as long-term trends allow us to rule out the interest rate hypothesis as well.

Eling and Kochanski (2013) provide a detailed overview of the extant literature on life insurance lapse. The aforementioned three hypotheses have been investigated by Schott (1971);

[1] For the remainder of this article we will use the terms lapse rate and hazard rate interchangeably.

Pesando (1974); Renshaw and Haberman (1986); Dar and Dodds (1989); Outreville (1990); Carson and Forster (2000); Kuo, Tsai and Chen (2003); Kim (2005); Cerchiara, Edwards and Gambini (2009); Milhaud, Loisel and Maume-Deschamps (2011); Kiesenbauer (2012); Russell et al. (2013), and Eling and Kiesenbauer (2013). Gottlieb and Smetters (2014) summarize the potential reasons to lapse as background risk and present a model where households and firms have heterogeneous priors thereof.

So far household level data has been used by Liebenberg, Carson and Dumm (2012); Fier and Liebenberg (2013); Belaygorod, Zardetto and Liu (2014), and Mulholland and Finke (2014). The use of microlevel variations in income represents a major step forward compared to previous studies that relied solely on macroeconomic variables in the aggregate, such as national unemployment rates. While Fier and Liebenberg (2013) and Mulholland and Finke (2014) use data from the Health and Retirement Study (HRS), which is a biannual survey, our data comes from a panel of ca. 20,000 households that is surveyed every quarter. The higher frequency of our data justifies the use of (continuous time) hazard rate models, which previously have been employed largely with administrative data obtained from a single insurer. Compared to the HRS, we also have more detailed contractual variables, notably the present duration as well as the identity of the insurer.

The rest of this paper is organized as follows. In Section 2, we describe our dataset and explain the sample construction process. In Section 3, we estimate our semiparametric baseline model, discuss the results and provide robustness tests. In Section 4, we extend the analysis to a fully parametric model, isolate marginal effects, and conduct a scenario analysis. In Section 5 we discuss motives other than the Emergency Fund Hypothesis as potential lapse drivers in our data. We conclude the analysis in Section 6. Additional technical descriptions and results are contained in the Appendix.

2. Data and Sample Construction

Our dataset is constructed from a representative household panel provided by a large German market research institute and covers the years 2005-2011 with a total of more than 44,000 households participating[2] . The average number of households participating in any given year

[2]To confirm the representativeness of our data we compare the rates of market participation in life insurance and other financial products from our data with the corresponding figures from the Eurosystem Household Finance and Consumption Survey. See Appendix C in Appendix C.

is above 20,000. Our final sample contains information on 34,301 insurance policies issued by 39 firms and held by 16,695 households. See Table 1 for descriptive statistics at the contract level.

[Table 1 about here]

Previous studies have documented substantial variation in lapse rates across different life insurance product types[3]. Liebenberg, Carson and Dumm (2012) distinguish between term life and cash-value policies, whereas Fier and Liebenberg (2013) cannot distinguish between different product types. Our data contains information on three separate types: Endowment life insurance, term life insurance, and unit-linked products. All other life insurance products are classified as "other insurance products". We focus our analysis on endowment life insurance, because it shows by far the deepest market penetration[4].

The data comes in five separate datasets: In the household dataset information on average monthly net income, location of residence, or the number of household members is stored for each household-year cell. The person dataset contains an individual's educational attainment, age, and role in the household (e.g. main income earner) in a given year. The inventory dataset contains all financial products (e.g. each checking account) that were held at the end of a given year as well as each product's year of inception and issuing firm. The acquisitions and terminations datasets represent the in- and outflow of the inventory dataset, respectively. They contain all financial products that were purchased or terminated in a given quarter. The terminations dataset furthermore provides information on the mode of termination, i.e. whether maturity was reached or whether a product was lapsed.

By merging the inventory with the terminations dataset, we obtain survival data for each endowment life insurance contract in our sample. For each contract we observe the year of inception and - if applicable - the quarter of termination. We define the onset of risk to be the fourth quarter of the year prior to the actual year of inception[5]. Table 2 provides an overview of the survival characteristics of our data, Figure 1 and Figure 2 show the survival and the hazard

[3] See Eling and Kiesenbauer (2013) for the German market.

[4] See Table 7 in Appendix C

[5] As the date of contract inception is recorded at annual frequency, but the date of termination is recorded at quarterly frequency, some convention is required. Using our definition, we overstate a contract's age by at most four quarters.

function, respectively. We merge the survival data with demographic characteristics from the household and person datasets. Using the inventory dataset, we construct an annual indicator for the ownership of other liquid financial assets, which we interpret as a proxy for wealth[6].

[Table 2 about here]

[Figure 1 about here]

[Figure 2 about here]

3. Evidence for the Emergency Fund Hypothesis

The methods we employ in this section are by now standard in the literature on lapse behavior. In our baseline regression we employ the Cox proportional hazards model, which has been used before by Pinquet, Guillén and Ayuso (2011) and Eling and Kiesenbauer (2013). As a robustness check, we run a pooled panel logit regression, which belongs to the class of generalized linear models. These have been used by Renshaw and Haberman (1986), Milhaud, Loisel and Maume-Deschamps (2011), and Fier and Liebenberg (2013) among others.

3.1. Preliminary Analysis

In a first step, we employ one-tailed two-proportion z-tests to show that the lapse rate is substantially higher for lower levels of income, wealth, and worse employment status. In each of the four tests we split the total sample of contracts into two subsamples. Then, we test the null hypothesis of an equal share of lapsed contracts in both subsamples against the alternative hypothesis that the share is lower in the subsample where wealth, income, or employment take less beneficial values. The four pairs of subsamples distinguish between contracts where the household owns real estate or not, where the household income is above the sample median or not, where the household holds liquid financial assets or not, and where the main income earner is unemployed or not. The test statistics are presented in Table 3. Each of the four null hypotheses is rejected at the 1%-level and in each of the four tests the share of lapsed contracts is more than 1 percentage point higher in the subsample with worse wealth, income, or em-

[6]We consider the following financial assets: Money market accounts, time deposit accounts, brokerage accounts, bonds, funds, and structured products. Current and savings accounts are excluded from liquid financial assets, as these are held by more than 80% of all households and therefore do not serve as an indicator for wealth.

ployment levels. Given that the overall share of lapsed contracts equals about 3 percent, this difference is substantial[7].

[Table 3 about here]

3.2. Baseline Estimation

After the univariate analysis we proceed to estimate a Cox proportional hazards model in order to examine the statistical significance of potential effects of income, wealth proxies, and employment status when other covariates are controlled for[8]. The model characterizes the hazard rate for each observation as the product of a common baseline hazard rate and an individual scaling factor depending on the covariate vector.

$$h(t_j|x_j) = h_0(t)exp(x_j'\beta) \tag{1}$$

where $h(t_j|x_j)$ = hazard rate for observation j,

$h_0(t)$ = common baseline hazard rate,

x_j = covariate vector of observation j.

The Cox proportional hazards model is semiparametric in the sense that no functional form is imposed on the baseline hazard rate $h_0(t)$, but it is estimated nonparametrically instead[9]. When - as is the case in our application - there is no theory available that suggests a particular form of the hazard rate the semiparametric approach has the virtue of being robust to misspecification errors. This comes at the cost that the economic significance of any potential effects is hard to assess. Quantifying the marginal effect of a change in a covariate on the survival function requires numerical integration of the estimated baseline hazard, which is impractical. Therefore, we will employ the Cox proportional hazards model in our baseline regression to examine statistical significance. Based on the results and the estimated shape of the baseline hazard, we will choose a fully parametric model that is analytically more tractable to evaluate the economic significance in Section 4.

[7] Though we observe households in the panel for seven consecutive years, it should be recalled that almost 93 percent of our observations are right censored given the long duration of life insurance contracts. This accounts for the low rate of lapses across observed contracts. (See, however, the estimated survival function in Figure 1.)

[8] We measure failures at quarterly frequency and yet employ a continuous-time framework in the estimation routine. As the average age at maturity is above 20 years and thereby more than 80 times greater than the length of one observation interval, this seems justified.

[9] The baseline hazard estimated from our data is depicted in Figure 2.

The following explanatory variables are included in the baseline regression: Household structure[10], age group of the policyholder at contract inception (nine ordinal levels), internet access, residence in the former GDR, urban or rural areas (three ordinal levels), the highest degree held by a household member (six ordinal levels), ownership of real estate, ownership of financial assets other than current or savings accounts, monthly net income (six ordinal levels)[11], the occupation of the main income earner (ten nominal groups), and calendar year dummies[12].

There may be unobservable firm characteristics that affect the hazard rate and that are correlated with wealth, income or employment status. We can partially account for this by stratifying our estimation by firms. Thereby we eliminate any time-invariant unobservable firm effect. The stratification procedure in hazard rate models resembles a fixed effects estimation in linear models, see Appendix D.1 for details.

The estimation results are presented in Table 4. Focusing on the emergency fund hypothesis (EFH), note first that income has the predicted negative effect on the hazard rate. Higher income is associated with a lower probability of lapse and the effect gets stronger as income increases. The hazard rate of policyholders in the highest income bracket is c.p. reduced by 51% relative to those in the lowest income bracket. We find a similar negative effect for our proxies of wealth. Owning real estate reduces the hazard rate c.p. by 19.9% and holding financial assets other than current or savings accounts reduces the hazard rate c.p. by 36.5%. Furthermore, we find significant effects for particular occupation groups: The hazard rate is c.p. lower when the household's main income earner is a white-collar worker in a leading management position or a pensioner rather than a white-collar worker in middle management. This may be attributed to the more stable income of these households or to otherwise unreported differences in income or wealth. When the main income earner is unemployed, the hazard rate is c.p. increased by 77.7%.

After controlling for income, in particular, education has no significant effect on lapse

[10]We distinguish between single males, single females, couples, and households with more than two members.

[11]The original dataset provides information on average monthly net income in 17 brackets of 250 Euro increments. In order to obtain sufficiently many observations for estimation in each bracket, we aggregate the 17 groups into six. The boundaries are chosen such as to allocate roughly 15% of the contract population into each group.

[12]In survival analysis models, the object of interest is the elapsed time since the onset of risk, the so-called *analysis time*. In our application, the onset of risk is given by the contract inception. Therefore, we implicitly control for cohort effects by including the calendar time, because each policy's date of inception (its cohort) is given by the calendar time minus the analysis time.

behavior[13]. Likewise, we find no significant effects for residence in East Germany, the household structure, rural or urban areas, or the age of the policyholder at contract inception. The latter result is worthwhile commenting on as it deviates from previous findings in the literature on lapse behavior. For instance, Cerchiara, Edwards and Gambini (2009) find lapse rates to decline steeply with policyholder age, whereas Eling and Kiesenbauer (2013) find a flat region for intermediate ages, which they attribute to more precise measurement of age. Both authors suggest this pattern may be driven by differing wealth levels and financial needs at different stages in the policyholder's life cycle. While Fier and Liebenberg (2013) still find significant effects of policyholder age despite including income and wealth measures, it should be noted that their dataset does not allow to disentangle the effects of contract age from those of policyholder age. In our analysis policyholder age has no effect on the lapse rate once we control both for income (as well as proxies of wealth) *and* contract age.

[Table 4 about here]

We accompany the baseline regression with a series of robustness checks. First, we estimate a pooled panel logit model on the quarterly lapse dummies[14]. As the baseline hazard exhibits non-monotonous time dependence, we include linear and quadratic terms of contract age[15]. Furthermore, we include firm dummies in addition to the explanatory variables used before[16]. The results are presented in Table 8 in Appendix C. Reassuringly, they are very similar to those from the hazard rate model.

4. Marginal Effects and Scenario Analysis

4.1. Model Selection and Robustness Check

In Section 3 we have employed the Cox proportional hazards model to show that income, wealth, and employment have statistically significant effects on the hazard rate, as is implied

[13]Outreville (2015) points out that a positive relation between the level of education and insurance demand found in cross-country studies may be due to multicollinearity between education, insurance demand, and other unobserved variables related to human economic development.

[14]The Cox proportional hazards model is a semi-parametric model of failure time. In contrast, the logistic regression model is a fully parametric model of the binary lapse variable. Moreover, in the logistic regression right-censoring is not accounted for. Annesi, Moreau and Lellouch (1989) prove that the Cox model is in general more efficient and discuss why regression results of these two models are oftentimes very similar.

[15]This is the main deviation of our specification from the one of Fier and Liebenberg (2013), as it allows us to separate the effects of contract age and policyholder age.

[16]Firm fixed effects are controlled for using dummies in the logit model and by stratifying the estimation in the hazard rate model.

by the Emergency Fund Hypothesis. In this section, we evaluate the economic significance of these effects. Specifically, we are interested in the discrete change of the survival function $S(t_j|x_j)$ due to a change in one of the variables related to the Emergency Fund Hypothesis. As it is impractical to obtain these marginal effects in semiparametric models, we introduce a fully parametric one.

To our knowledge, parametric models for survival data have not yet been employed in the context of life insurance lapse. We choose the loglogistic model of the class of accelerated failure time models[17]. This model has three desirable properties. First, it exhibits the smallest Akaike information criterion when compared to other candidates, see Table 9 in Appendix C. In addition, there is a closed-form expression for the marginal effect of a change in the covariates on the survival function when using the loglogistic specification[18]. Most importantly however, the model allows for both, a monotonically decreasing as well as a hump-shaped hazard rate. The nonparametric estimate of the hazard rate depicted in Figure 2 suggests that it may have a hump-shaped form. Using the loglogistc model, we are able to test this hypothesis at the firm level.

The results from the loglogistic regression are presented in Table 10 in Appendix C. Reassuringly, they are qualitatively similar to the results from the Cox proportional hazards regression. To see this, note that the coefficients for income, wealth proxies, and occupational statuses exhibit similar statistical significance levels as in the baseline regression. In addition, the statistically significant coefficients have opposite signs to their counterparts in the baseline regression, which resembles similar economic effects[19]. We test the null hypothesis of a monotonically decreasing hazard rate for each of the 39 firms[20]. The null hypothesis is rejected in favor of a unimodal hazard rate for three firms at the 5%-level and for two firms at the 10%-level. However, the joint hypothesis of monotonically decreasing hazard rates for all firms is rejected at the 5% level. We conclude that the hump-shaped form of the hazard rate suggested by Figure 2 is driven by a few number of firms and that for most firms the hazard rate can be

[17]See Kalbfleisch and Prentice (2002, ch. 7) or Cleves et al. (2010, ch. 12-13) for an introduction to the accelerated failure time formulation of survival analysis models.

[18]We explicitly state our measure of marginal effects in Appendix D.2.

[19]A lower probability of failure in a given period of time corresponds with negative coefficients in the Cox proportional hazards regression and with positive coefficients in the class of accelerated failure time models. See Cleves et al. (2010, p. 240) for details.

[20]See the test statistics in Table 11 in Appendix C.

assumed to be monotonically decreasing.

4.2. Marginal Effects

In the Cox proportional hazards model a change in the covariates leads to a proportional shift of the hazard rate, such that the effect can be expressed in terms of a hazard ratio that is time-invariant. In general, accelerated failure time models do not have this property. Marginal effects on the survival function vary by analysis time. Therefore, we tabulate marginal effects at 0, 5, 10, 15, and 20 years into the contract's life. In addition, differences in the covariate vector that persist for a longer period of time also cause larger differences in the survival function[21]. We consider the difference in survival functions one year, i.e. four periods of analysis time, after the change in covariates comes into effect. Finally, we consider marginal effects for the median contract. Specifically, we fix the ordinal covariates at the sample median and the nominal ones at the sample mode[22] during the period $(0,t']$. We then consider the difference in survival probabilities resulting from a detrimental state of the covariate in question vis-à-vis a continuation of the median state during the period $(t',t'+4]$. t' denotes the time at which the change in covariates occurs[23].

The results are presented in Table 5. As we focus on the Emergency Fund Hypothesis, we tabulate the marginal effects for income, our two wealth proxies, and employment status. First, note that in terms of absolute value the marginal effects are highest when the shock occurs five years after origination of the contract. Shocks occurring at later stages of the contract's life have smaller effects on the survival function in absolute terms. This result is due to the decreasing hazard rate found earlier[24]. Among the four variables considered, the effect of unemployment is the largest. When the main income earner experiences a transition from

[21]When compared to the default case of continuous employment, the probability of surviving one year of unemployment is different from the probability of surviving two years of unemployment simply because in the latter case the unemployment effect is operational for a longer period of time.

[22]Note that we also set the firm dummies to their sample mode, i.e. consider the firm with the highest market share during 2005-2011.

[23]The marginal effect mechanically depends on both, analysis time as well as the history of covariates. Therefore, averaging marginal effects over the sample is not an option. We consider marginal effects for the sample median instead.

[24]For firm #1 in Table 11 in Appendix C, which is the default firm we consider, the estimate of the hazard rate's shape parameters is numerically negative. This resembles a hump-shaped hazard rate and therefore marginal effects right at the origination of the contract are smaller in absolute value than five years later. However, the hypothesis of a monotonically decreasing hazard rate is not rejected at common levels for the default firm. Likewise, the difference between marginal effects right at the beginning of a contract's life and those five year later is not statistically significant either.

the mode employment status (a white-collar job in middle management) to unemployment the survival probability decreases by up to 0.79 percentage points per year. A reduction of monthly net household income from the median to the lowest category decreases the survival function by up to 0.18 percentage points per year, however this figure is not statistically significant. When we consider a reduction from the top to the bottom income category we find a statistically significant reduction of the survival function by up to 0.46 percentage points per year. A change in the wealth proxies can be interpreted as a negative shock to the household's net worth. The median household owns real estate and holds financial assets other than current or savings accounts. Decreasing holdings in real estate and financial assets reduces the survival function by up to 0.22 and 0.41 percentage points per year, respectively.

[Table 5 about here]

4.3. Compound Effects

In Section 4.2 we have considered the isolated effects of a deterioration of income levels, wealth proxies, and employment status. Keep in mind that these marginal effects are net of any other changes in the covariate vector. However, in practice they often accompany each other. In order to link our results to the Emergency Fund Hypothesis and gain a more realistic picture of the effect of unemployment, we conduct a scenario analysis. Specifically, we compare the survival functions in two scenarios that realistically combine the marginal effects quantified before. In addition, we show how these effects accumulate over time. In both scenarios we consider a contract that is originated on 1st January 2005 and compute end-of-year survival probabilities for $2005-2011$. In the baseline scenario we assume the respective household's main income earner has a white-collar job in middle management, which equals the sample mode employment status. For the remaining covariates we assume the sample median values *conditional* on middle white-collar employment status[25]. In particular, monthly household income is assumed to be in the $2,000-2,499$ Euro bracket between 2005 and 2008 and in the $2,500-2,999$ Euro bracket as of 2009. Moreover, the household is assumed to own real estate. In the alternative scenario, we assume the household's main income earner is unemployed. Likewise, we assume the remaining covariates take the sample median values *conditional* on unemployment.

[25]When computing marginal effects in Section 4.2, we set the covariate vector to the unconditional sample median/mode.

Monthly household income is in the $0-1,499$ Euro bracket for all years and the household does not own real estate[26]. However, in both scenarios the household holds liquid financial assets other than current and savings accounts. The survival probabilities resulting from both scenarios are plotted in Figure 3 and the corresponding numbers including significance tests are presented in Table 12 in Appendix C. As of 2006 the difference between the two survival probabilities is significant at the 5%-level and it accumulates to over 9 percentage points over the course of seven years.

[Figure 3 about here]

5. Discussion of Other Hypotheses

Recall from the Introduction that the literature has presented two alternative hypotheses to account for observed lapse behavior. According to the policy replacement hypothesis (PRH), policyholders may want to switch to a more attractive competing contract. Alternatively, they may want to invest their savings elsewhere when this becomes more attractive, notably when market interest rates are relatively high compared to the expected return from continuation of the insurance contract (interest rate hypothesis, IRH).

We first present evidence that replacements, as envisaged by the PRH, occur only seldom and that income and wealth effects are robust to their exclusion. The panel nature of our dataset allows us to track policyholders even after the lapse and to examine whether they or other household members establish a new life insurance with another firm. For every one of the lapses in our dataset we construct a dummy variable that equals unity when the lapse occurs in the four quarters centered around the acquisition of another endowment life insurance product by the respective household. Thereby we apply a conservative notion of replacement, as we do not require the same person to buy a new life insurance, but take into account new contracts signed by anyone of the household members. Using this very broad definition of replacement, it is satisfied by only 55 out of the observed 1001 lapses, i.e. the overall replacement rate is 5.49%.

[26]The different real estate holdings in the two subsamples reflect that households who are less wealthy are generally at a higher risk of becoming unemployed. For instance, lagged income and wealth levels of households whose main income earner became unemployed during the last year are significantly lower than the sample median.

It is notable that Fier and Liebenberg (2013) find a much higher replacement rate of 13.7%. Various factors may explain this difference. Due to the biannual frequency of the survey data, Fier and Liebenberg (2013) have to consider all lapses in the two years prior to a new acquisition instead of a narrower frame around the acquisition. If we apply their definition, the replacement rate is, however, still only 6.09% in our dataset. Importantly, they make no distinction between different product types, whereas we restrict ourselves to endowment life insurance contracts. As a final robustness test towards the policy replacement hypothesis, we restrict the sample to those policies where no acquisition of another policy by the same household is recorded throughout the observation period. Then we re-estimate the Cox proportional hazards model as well as the pooled panel logit. All results remain qualitatively similar to the unrestricted case, see Table 13 in Appendix C.

Finally, we argue that from 2005 to 2011, which is the time span that we consider, it would have been generally unattractive for policyholders to lapse in order to pursue an alternative investment possibility. German insurers' (smoothed) capital gains, which ultimately are paid out to policyholders over time, were notably above 4%.[27] This exceeded, in particular, even the long-term (risk-free) interest rate in this period. Moreover, Outreville (1990) and Kuo, Tsai and Chen (2003) both find that interest rates have no short-term effect but affect aggregate lapse rates only in the long-run. Since also the long-run trend of interest rates has been declining since Reunification, we conclude that the interest rate hypothesis should not be a main driver of lapses, at least not during our period of observation.

A further motive for lapse arises from reclassification risk. When policy premiums depend on the policyholder's health status, a deteriorated health status may make a continuation of the policy unaffordable[28]. Hendel and Lizzeri (2003) formalize this notion. Fang and Kung (2012) estimate a model of life insurance demand using data from the Health and Retirement Study (HRS). They find health shocks to be an important driver of life insurance lapse among elderly policyholders. Since the pricing of endowment life insurance policies is not based on the policyholder's health status in Germany, we argue that reclassification risk does not pose a motive to lapse in our application.

[27] See Wichert (2013).

[28] We thank J. François Outreville for bringing this rationale to our attention.

6. Conclusion

In this paper we re-examine the emergency fund hypothesis for the reason to lapse a life insurance. Methodologically, we follow Cerchiara, Edwards and Gambini (2009) and Eling and Kiesenbauer (2013) in estimating proportional hazard rate models. In addition, we quantify effects on the survival function using a fully parametric model. The emergency fund hypothesis has previously been investigated by Outreville (1990) and Kuo, Tsai and Chen (2003) using aggregate data from the United States and Canada. Liebenberg, Carson and Dumm (2012) and Fier and Liebenberg (2013) are the only other studies using microlevel data in the context of the EFH. Applying generalized linear models, they find evidence for the emergency fund hypothesis in the US market. The German market has previously been studied by Kiesenbauer (2012) and Eling and Kiesenbauer (2013) using administrative data. We are the first study to combine survival analysis techniques with microlevel information and provide further evidence in favor of the emergency fund hypothesis. Moreover, we are the first study to employ parametric models of survival data in the context of life insurance lapse.

We restrict our analysis to endowment life insurance, as this product exhibits the highest market penetration among all life insurance products in Germany and plays an important role for saving and retirement planning. The main three hypotheses discussed in the literature so far, the emergency fund hypothesis, the policy replacement hypothesis, and the interest rate hypothesis are not mutually exclusive. However, we can rule out the lapse motive as put forward by the policy replacement hypothesis for endowment life insurance because of the high loss incurred from a lapse relative to serving the contract till maturity. Furthermore, we argue that because of the market conditions during our period of observation the interest rate hypothesis is unlikely to motivate lapses either. This strengthens our evidence for the emergency fund hypothesis.

The novel features of our dataset, most notably microlevel information on income and wealth as well as contract age and firm, allow us to quantify and partially refute previously established findings. Our key results are that policyholder age has no effect on the lapse rate once income, wealth, and contract age are controlled for and that unemployment poses a severe threat to policy survival. The former result contradicts all previous microlevel studies. The latter is in line with Outreville (1990) and Liebenberg, Carson and Dumm (2012), but contradicts the finding of Kuo, Tsai and Chen (2003) that unemployment plays an economically insignificant role. Specifically, we find transition to unemployment to increases the lapse rate by more than 75%. Regarding stress tests that link a deterioration of macroeconomic conditions to a mass lapse scenario, the discrepancy between our results and those of Kuo, Tsai and Chen (2003) highlights that selection into the life insurance market needs to be taken into account.

The main results of our paper are found using information other than policyholder's age, gender, and location of residence, which are typically the only data German insurance firms may collect about their customers. Nevertheless, our results have practical implications for insurers' internal pricing models. For instance, firms may employ regional unemployment figures to forecast lapse rates for specific customer segments, rather than relying on economy-wide numbers. Likewise, information on wealth can be extrapolated from geographical information. Insurance companies considering market entry or already operating in emerging market economies may consider the following aspect of our findings. Our results suggest that a population's demographic profile has little effect on aggregate lapse rates once income and wealth are taken into account.

We show that unemployment has a substantial impact on the lapse rate based on data from Germany. The extent to which these results translate to other economies with different labor market institutions remains an open question left for future research.

References

Annesi, Isabella, Thierry Moreau, and Joseph Lellouch. 1989. "Efficiency of the logistic regression and cox proportional hazards models in longitudinal studies." *Statistics in Medicine*, 8(12): 1515–1521.

Belaygorod, Anatoliy, Atilio Zardetto, and Yuanjin Liu. 2014. "Bayesian Modeling of Shock Lapse Rates Provides New Evidence for Emergency Fund Hypothesis." *North American Actuarial Journal*, 18(4): 501–514.

Carson, James M., and Mark D. Forster. 2000. "Suitability and Life Insurance Policy Replacement." *Journal of Insurance Regulation*, 18: 427–447.

Cerchiara, Rocco R., Matthew Edwards, and Alessandra Gambini. 2009. "Generalized Linear Models in Life Insurance: Decrements and Risk Factor Analysis under Solvency II." *Giornale dell'Istituto Italiano degli Attuari*, 100–122.

Cleves, Mario, William W. Gould, Roberto G. Gutierrez, and Yulia Marchenko. 2010. *An Introduction to Survival Analysis Using Stata. Stata Press books*, StataCorp LP.

Dar, A., and C. Dodds. 1989. "Interest Rates, the Emergency Fund Hypothesis and Saving through Endowment Policies: Some Empirical Evidence for the U.K." *Journal of Risk and Insurance*, 56(3): pp. 415–433.

Eling, Martin, and Dieter Kiesenbauer. 2013. "What Policy Features Determine Life Insurance Lapse? An Analysis of the German Market." *Journal of Risk and Insurance*.

Eling, Martin, and Michael Kochanski. 2013. "Research on lapse in life insurance: what has been done and what needs to be done?" *Journal of Risk Finance*, 14(4): 392–413.

Eurosystem Household Finance and Consumption Network. 2013. "The Eurosystem Household Finance and Consumption Survey - Results from the first wave." European Central Bank Statistics Paper Series 2.

Fang, Hanming, and Edward Kung. 2012. "Why Do Life Insurance Policyholders Lapse? The Roles of Income, Health and Bequest Motive Shocks." National Bureau of Economic Research Working Paper 17899.

Fier, Stephen G., and Andre P. Liebenberg. 2013. "Life Insurance Lapse Behavior." *North American Actuarial Journal*, 17(2): 153–167.

Gottlieb, Daniel, and Kent Smetters. 2014. "Lapse-Based Insurance." The Wharton School, The University of Pennsylvania working paper.

Hendel, Igal, and Alessandro Lizzeri. 2003. "The Role of Commitment in Dynamic Contracts: Evidence from Life Insurance." *The Quarterly Journal of Economics*, 118(1): pp. 299–327.

Kalbfleisch, John D., and Ross L. Prentice. 2002. *The Statistical Analysis of Failure Time Data. Wiley Series in Probability and Statistics*. 2 ed., Hoboken, New Jersey:John Wiley & Sons, Inc.

Kiesenbauer, Dieter. 2012. "Main Determinants of Lapse in the German Life Insurance Industry." *North American Actuarial Journal*, 16(1): 52–73.

Kim, Changki. 2005. "Modeling Surrender and Lapse Rates With Economic Variables." *North American Actuarial Journal*, 9(4): 56–70.

Kuo, Weiyu, Chenghsien Tsai, and Wei-Kuang Chen. 2003. "An Empirical Study on the Lapse Rate: The Cointegration Approach." *Journal of Risk and Insurance*, 70(3): 489–508.

Liebenberg, Andre P., James M. Carson, and Randy E. Dumm. 2012. "A Dynamic Analysis of the Demand for Life Insurance." *Journal of Risk and Insurance*, 79(3): 619–644.

Milhaud, Xavier, Stéphane Loisel, and Véronique Maume-Deschamps. 2011. "Surrender triggers in Life Insurance: what main features affect the surrender behavior in a classical economic context?" *Bulletin Français d'Actuariat*, 22: 5–48.

Mulholland, Barry S., and Michael S. Finke. 2014. "Does Cognitive Ability Impact Life Insurance Policy Lapsation?" Available at SSRN: http://ssrn.com/abstract=2512076.

Outreville, J. Francois. 1990. "Whole-life insurance lapse rates and the emergency fund hypothesis." *Insurance: Mathematics and Economics*, 9(4): 249–255.

Outreville, J. Francois. 2015. “The Relationship between Relative Risk Aversion and the Level of Education: A Survey and Implications for the Demand for Life Insurance.” *Journal of Economic Surveys*, 29(1): 97–111.

Pesando, James E. 1974. “The Interest Sensitivity of the Flow of Funds Through Life Insurance Companies: An Econometric Analysis.” *Journal of Finance*, 29(4): pp. 1105–1121.

Pinquet, Jean, Montserrat Guillén, and Mercedes Ayuso. 2011. “Commitment and Lapse Behavior in Long-Term Insurance: A Case Study.” *Journal of Risk and Insurance*, 78(4): 983–1002.

Renshaw, A. E., and S. Haberman. 1986. “Statistical Analysis of Life Assurance Lapses.” *Journal of the Institute of Actuaries*, 113(3): 459–497.

Russell, David, Stephen G. Fier, James M. Carson, and Randy E. Dumm. 2013. “An Empirical Analysis of Life Insurance Policy Surrender Activity.” *Journal of Insurance Issues*, 36(1): 35–57.

Schott, Francis H. 1971. “Disintermediation Through Policy Loans at Life Insurance Companies.” *Journal of Finance*, 26(3): pp. 719–729.

Wichert, Björn. 2013. “Überschussbeteiligungen sinken auf breiter Front.” VersicherungsJournal Verlag GmbH VERS117593.

A. Tables for the Main Body

Table 1: Policyholder characteristics

	Mean	Median	Interquartile range
Monthly real net income of household[1]	2,469.41	2,400.63	1358.44
Age of policyholder at inception[2]	32.76	31.00	18.00
Age of policy[2]	16.60	15.50	11.50
Highest degree of policyholder[3]		344	353-7
Policyholder is female[4]	44.50		
Household owns savings account[4]	73.13		
Household owns other liquid financial assets[4,5]	65.51		
Household owns real estate[4]	62.26		

Descriptive sample statistics based on $34,301$ contracts.
[1] In June 2010 prices, EUR
[2] In years
[3] Using the International Standard Classification of Education (ISCED). 344: General upper secondary education with direct access to tertiary education. 353: Vocational upper secondary education without direct access to tertiary education. 7: Master's or equivalent.
[4] Dummy variable in percentage points
[5] Liquid financial assets include: Money market accounts, time deposit accounts, brokerage accounts, bonds, funds, and structured products.

Table 2: Survival characteristics at the contract level

	N	Mean	Median	Interquartile range
Left-truncated[1]	34,301	97.84		
Right-censored[1]	34,301	92.98		
Matured[1]	34,301	3.78		
Lapsed[1]	34,301	2.92		
Age at first observation[2]	34,301	14.04	13.00	12.00
Age at last observation[2]	34,301	17.80	17.00	12.00
Age at maturity[2]	1,298	21.58	20.00	14.00
Age at lapse[2]	1,001	14.16	13.50	9.50

[1] Dummy variable in percentage points

[2] In years

Table 3: Lapse rate by income, wealth, and employment

Covariate	Subsample	N	Lapse rate (%)	Diff	z-Stat	P-value
Income	$\leq$ median	18,916	3.68	1.70	9.2865	0.0000
	$>$ median	15,385	1.98			
Real Estate	0	12,602	3.71	1.26	6.6698	0.0000
	1	21,699	2.46			
Fin. assets	0	11,563	4.04	1.69	8.7918	0.0000
	1	22,738	2.35			
Employment	0	582	8.76	5.95	8.4489	0.0000
	1	33,719	2.82			

One-tailed two-proportion z-tests. The sample of $34,301$ contracts is split into two subsamples according to a criterion specified in the first two columns. The share of lapsed contracts is calculated for each subsample and the difference in lapse rates is considered. The p-values refer to the null hypothesis of a zero difference.

Table 4: Baseline model

Covariate	Level	Coefficient	Std. err.	Haz. ratio
Nominal monthly net income of household (EUR)	(0-1499)			
	1500-1999	-0.191	0.132	0.826
	2000-2499	-0.216	0.132	0.806
	2500-2999	-0.261*	0.148	0.771
	3000-3999	-0.447***	0.150	0.640
	4000+	-0.714***	0.182	0.490
Wealth	Real estate ownership	-0.212**	0.089	0.809
	Financial assets	-0.455***	0.084	0.635
Occupation of main income earner	Retired	-0.279*	0.160	0.756
	In education/not in labor force	-0.021	0.321	0.979
	Low blue-collar	-0.250	0.258	0.779
	High blue-collar	-0.008	0.125	0.992
	Low white-collar	-0.034	0.173	0.967
	(Middle white-collar)			
	High white-collar	-0.323*	0.194	0.724
	State-employed	0.097	0.147	1.102
	Self-employed	0.197	0.162	1.217
	Unemployed	0.575***	0.198	1.777
Highest degree in household	No vocational training	0.514**	0.249	1.673
	Lower secondary education	0.075	0.143	1.078
	Middle secondary education	0.094	0.104	1.099
	Advanced secondary education	-0.233	0.171	0.792
	Postsecondary education	-0.101	0.120	0.904
	(University degree)			
Year	(2005)			
	2006	-0.067	0.130	0.935
	2007	-0.155	0.131	0.856
	2008	-0.255*	0.140	0.775
	2009	-0.334**	0.149	0.716
	2010	-0.438***	0.157	0.645
	2011	-0.706***	0.163	0.494
No internet access		-0.221**	0.108	0.801
Residence in East Germany		0.111	0.097	1.118
Age of policyholder at inception (years)	0-17	-0.147	0.168	0.864
	18-22	0.062	0.131	1.064
	23-25	0.124	0.140	1.132
	(26-30)			
	31-35	0.042	0.129	1.043
	36-40	0.030	0.137	1.030
	41-45	0.051	0.148	1.052
	46-55	0.057	0.165	1.059
	56+	-0.079	0.222	0.924
Household structure	Single male	-0.162	0.159	0.851
	Single female	-0.172	0.157	0.842
	(Couple)			
	More than two persons	0.057	0.094	1.058
Town size	(0-20 thousand)			
	20-200 thousand	-0.047	0.091	0.954
	>200 thousand	-0.085	0.107	0.919

Cox proportional hazards regression. 116,201 observations; 34,301 contracts; 1,001 failures; 461,278 quarters at risk. Standard errors adjusted for 16,695 household clusters. Estimation stratified by 39 firms. $^{*}p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$. Reference levels in parentheses. Wealth is proxied by two dummy variables that indicate whether a household owns real estate and whether a household holds any financial assets other than a current or a savings account. Positive coefficients indicate an increased hazard rate. The hazard ratio indicates the factor by which the hazard rate changes in response to a change in the covariate relative to its reference level. For instance, an increase in monthly net income from the 0-1499 bracket to the 1500-1999 bracket will reduce the hazard rate c.p. by 17.4%.

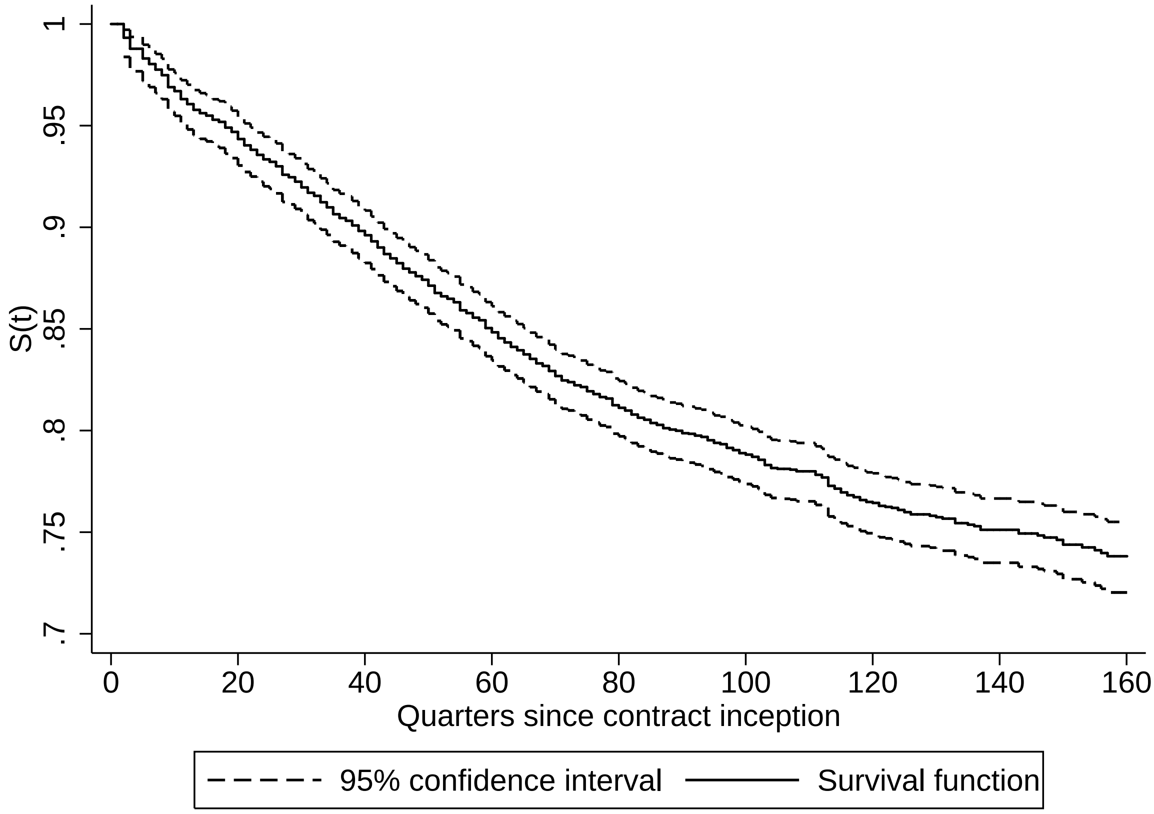
Table 5: Change in the survival function

Covariate	Shock period $(t', t'+4]$	years	Estimated change	Std. err.	P-value
Employment status	$(0,4]$	0-1	-0.767	0.514	0.136
	$(20,24]$	5-6	-0.788	0.406	0.053
Mode: White collar in middle position Shock: Unemployment	$(40,44]$	10-11	-0.718	0.365	0.049
	$(60,64]$	15-16	-0.647	0.326	0.048
	$(80,84]$	20-21	-0.581	0.291	0.046
Income	$(0,4]$	0-1	-0.174	0.147	0.237
	$(20,24]$	5-6	-0.184	0.121	0.128
Median: 2000-2499 Shock: 0-1499	$(40,44]$	10-11	-0.173	0.110	0.117
	$(60,64]$	15-16	-0.160	0.101	0.113
	$(80,84]$	20-21	-0.147	0.093	0.111
Real estate ownership	$(0,4]$	0-1	-0.203	0.131	0.120
	$(20,24]$	5-6	-0.215	0.098	0.028
Median: Yes Shock: No	$(40,44]$	10-11	-0.202	0.091	0.027
	$(60,64]$	15-16	-0.186	0.085	0.028
	$(80,84]$	20-21	-0.172	0.079	0.030
Financial assets	$(0,4]$	0-1	-0.393	0.216	0.069
	$(20,24]$	5-6	-0.411	0.131	0.002
Median: Yes Shock: No	$(40,44]$	10-11	-0.382	0.120	0.001
	$(60,64]$	15-16	-0.350	0.112	0.002
	$(80,84]$	20-21	-0.319	0.104	0.002

Estimated reduction of the survival function in a "shock" scenario. In the reference scenario all covariates are held fixed at the sample median (ordinal covariates) or at the sample mode (nominal covariates) during the interval $(0,t'+4]$ of analysis time. Four periods of analysis time resemble one year of calendar time (quarterly data). In the shock scenario the given covariate switches to the indicated shock-status during the interval $(t',t'+4]$. That is, after $t'/4$ "normal" years there is one "bad" year. In both scenarios all other covariates are fixed at the sample median/mode throughout $(0,t'+4]$. The calendar year is fixed at 2007 for the entire lifetime of the contract, firm dummies are fixed at the sample mode (firm with highest market share). The change in the survival function at time $t'+4$ resulting from the shock is given in percentage points. The estimation is based on the loglogistic regression (see Table 10 in Appendix C.). The p-values refer to the null hypothesis of a zero reduction.

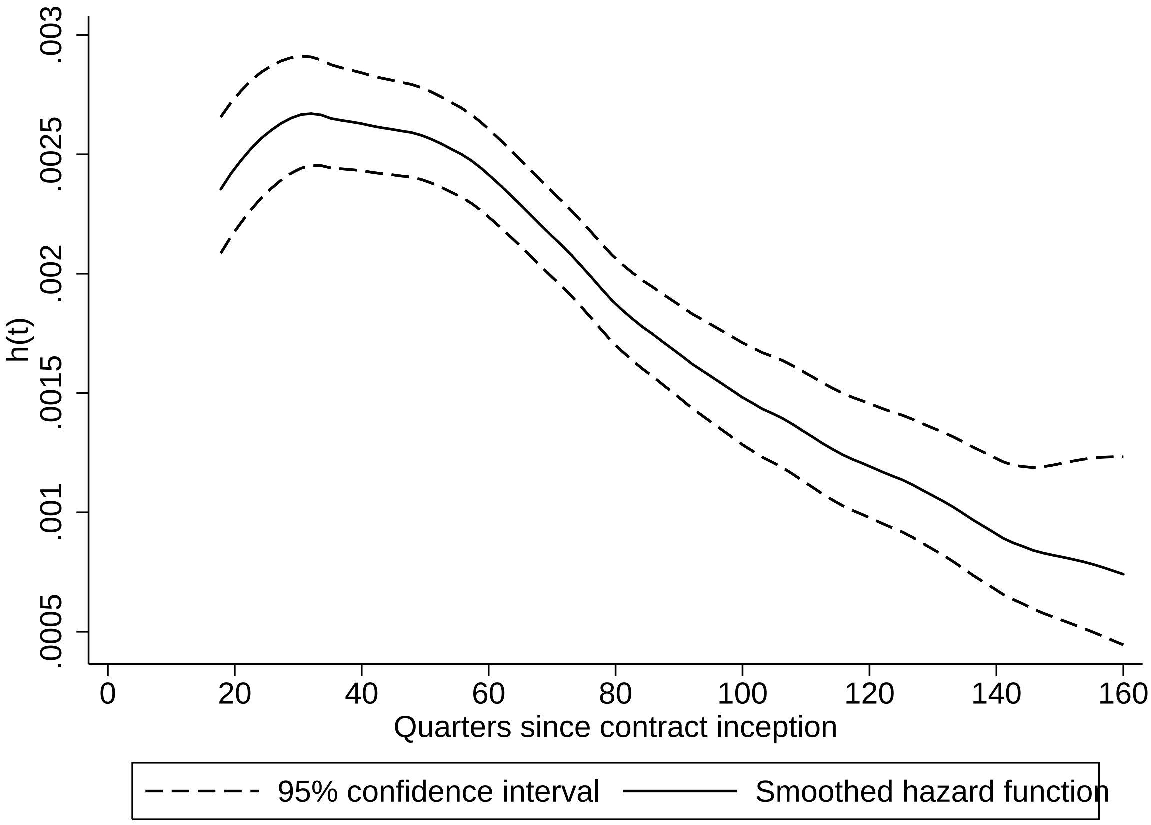
B. Figures for the Main Body

Figure 1: Survival function



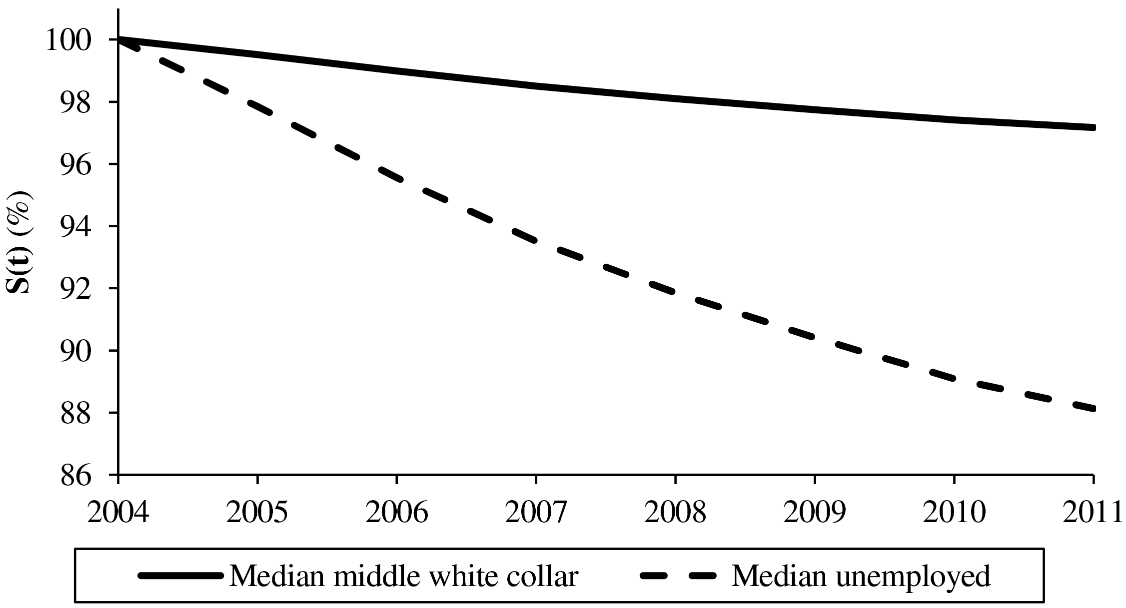
Kaplan-Meier estimate of the survival function based on the $34,301$ contracts in our sample.

Figure 2: Hazard function



Estimated hazard function based on the 34, 301 contracts in our sample.

Figure 3: Survival function given employment status



Estimated survival function for a contract originated on 1st January 2005 in two different scenarios. Reference scenario: All covariates fixed at the sample median/mode *conditional* on the main income earner working a white-collar job in middle management. Alternative scenario: All covariates fixed at the sample median/mode *conditional* on the main income earner being unemployed. See Table 12 in Appendix C for corresponding numbers and significance tests.

C. Tables for a Separate Appendix

Table 6: Comparison of our data with the HFCS[1]

Product	Participation rate from the HFCS	Participation rate from our data
Ownership of household main residence	44.2	49.9
Deposits	99.0	99.5
Shares[2]	10.6	19.2
Mutual Funds[2]	16.9	
Bonds	5.2	8.2
Voluntary private pension/whole life insurance[3]	46.5	46.0

Entries represent the share of households participating in the respective market in percentage points. $16,695$ unique households, $58,652$ observations.

[1] External sources (HFCS): Eurosystem Household Finance and Consumption Network (2013, Table 2.1, p.24 and Table 2.4, p.37)

[2] Our data provides information on ownership of brokerage accounts, rather than shares. We contrast ownership of brokerage accounts with shares as well as participation in mutual funds, because the latter requires a brokerage account as well.

[3] We report participation in any of the following products: Personal pension schemes, pension funds, long-term care insurance, and total permanent disability insurance.

Table 7: Market penetration of life insurance types

Year	Participating households	Endowment	Term life	Unit-linked
2005	20,680	43.25	13.45	9.38
2006	21,256	42.67	14.02	9.32
2007	21,834	41.18	14.57	8.56
2008	21,054	39.56	14.62	8.11
2009	20,188	37.93	13.99	7.77
2010	20,108	36.75	13.71	7.38
2011	20,894	35.49	13.98	6.82

Entries represent the share of households in our data holding the respective product in percentage points. $16,695$ unique households, $146,014$ observations.

Table 8: Robustness check for baseline regression

Covariate	Level	Coeff.	Std. err.
Contract age (quarters)	linear	-0.001	0.004
	quadratic	-0.000*	0.000
Nominal monthly net income of household (EUR)	(0-1499)		
	1500-1999	-0.200	0.133
	2000-2499	-0.237*	0.132
	2500-2999	-0.270*	0.149
	3000-3999	-0.478***	0.150
	4000+	-0.737***	0.181
Wealth	Real estate ownership	-0.207**	0.090
	Financial assets	-0.466***	0.084
Occupation of main income earner	Retired	-0.218	0.161
	In education/not in labor force	-0.015	0.329
	Low blue-collar	-0.317	0.258
	High blue-collar	-0.046	0.127
	Low white-collar	-0.074	0.176
	(Middle white-collar)		
	High white-collar	-0.356*	0.195
	State-employed	0.118	0.149
	Self-employed	0.202	0.165
	Unemployed	0.573***	0.199
Highest degree in household	No vocational training	0.487*	0.250
	Lower secondary education	0.113	0.144
	Middle secondary education	0.101	0.105
	Advanced secondary education	-0.285	0.174
	Postsecondary education	-0.093	0.120
	(University degree)		
Year	(2005)		
	2006	-0.070	0.131
	2007	-0.133	0.129
	2008	-0.237*	0.139
	2009	-0.332**	0.147
	2010	-0.395**	0.154
	2011	-0.699***	0.162
No internet access		-0.215**	0.109
Residence in East Germany		0.110	0.097
Age of policyholder at inception (years)	0-17	-0.179	0.169
	18-22	0.086	0.132
	23-25	0.147	0.140
	(26-30)		
	31-35	0.075	0.130
	36-40	0.035	0.138
	41-45	0.038	0.146
	46-55	0.032	0.166
	56+	-0.139	0.215
Household structure	Single male	-0.178	0.159
	Single female	-0.167	0.158
	(Couple)		
	More than two persons	0.082	0.095
Town size	(0-20 thousand)		
	20-200 thousand	-0.075	0.092
	>200 thousand	-0.110	0.109
Constant		-3.683***	0.260

Pooled panel logit regression. The dependent variable is a dummy taking value one when a contract was lapsed in a given year. Lapsed contracts are dropped from the sample in the following years. $116,201$ observations. ${}^{*}p < 0.1$; ${}^{**}p < 0.05$; ${}^{***}p < 0.01$. Standard errors adjusted for $16,695$ household clusters. Fixed effects of 39 firms are not shown. Reference levels in parentheses. Wealth is proxied by two dummy variables that indicate whether a household owns real estate and whether a household holds any financial assets other than a current or a savings account. Positive coefficients indicate an increased probability of lapse.

Table 9: Selection of the parametric model

Distribution of the error term	Possible features of the hazard rate: time-varying	non-monotonous	Number of parameters	AIC
Exponential	no	no	81	6,322.568
Weibull	yes	no	120	6,372.004
Gompertz	yes	no	120	6,359.490
Lognormal	yes	yes	120	6,342.773
Loglogistic	yes	yes	120	6,342.462

Accelerated failure time models by underlying distribution. $34,301$ contracts; $1,001$ failures; $461,278$ quarters at risk. Estimation stratified by 39 firms. Entries in the column titled AIC refer to the Akaike information criterion.

Table 10: Parametric model

Covariate	Level	Coefficient	Std. err.
Nominal monthly net income of household (EUR)	(0-1499)		
	1500-1999	0.204	0.159
	2000-2499	0.269*	0.152
	2500-2999	0.280	0.176
	3000-3999	0.549***	0.174
	4000+	0.852***	0.202
Wealth	Real estate ownership	0.310***	0.112
	Financial assets	0.530***	0.099
Occupation of main income earner	Retired	0.432***	0.162
	In education/not in labor force	0.127	0.356
	Low blue-collar	0.321	0.279
	High blue-collar	0.002	0.144
	Low white-collar	0.004	0.235
	(Middle white-collar)		
	High white-collar	0.374*	0.212
	State-employed	-0.150	0.193
	Self-employed	-0.193	0.196
	Unemployed	-0.857***	0.294
Highest degree in household	No vocational training	-0.540**	0.275
	Lower secondary education	-0.071	0.167
	Middle secondary education	-0.130	0.128
	Advanced secondary education	0.235	0.209
	Postsecondary education	0.140	0.132
	(University degree)		
Year	(2005)		
	2006	-0.023	0.156
	2007	0.074	0.148
	2008	0.261*	0.153
	2009	0.376***	0.159
	2010	0.458***	0.185
	2011	0.759***	0.180
No internet access		0.302**	0.137
Residence in East Germany		-0.151	0.119
Age of policyholder at inception (years)	0-17	0.301	0.189
	18-22	-0.120	0.177
	23-25	-0.196	0.171
	(26-30)		
	31-35	-0.081	0.140
	36-40	-0.171	0.161
	41-45	-0.234	0.166
	46-55	-0.316*	0.188
	56+	-0.255	0.222
Household structure	Single male	0.176	0.179
	Single female	0.109	0.191
	(Couple)		
	More than two persons	-0.172	0.114
Town size	(0-20 thousand)		
	20-200 thousand	0.092	0.104
	>200 thousand	0.136	0.126
Constant		4.572***	0.353

Loglogistic regression. 116,201 observations; 34,301 contracts; 1,001 failures; 461,278 quarters at risk. Standard errors adjusted for 16,695 household clusters. Estimation stratified by 39 firms; firm coefficients omitted. $^{*}p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$. Reference levels in parentheses. Wealth is proxied by two dummy variables that indicate whether a household owns real estate and whether a household holds any financial assets other than a current or a savings account. Positive coefficients indicate an increased probability of survival, i.e. a decreased probability of lapse.

Table 11: Tests of a monotonically decreasing hazard rate

Firm ID	Obs.	$\widehat{\ln(\gamma)}$	P-value	Firm ID	Obs.	$\widehat{\ln(\gamma)}$	P-value
1	17,610	-0.056	0.6926	21	4,028	-0.010	0.9746
2	1,030	-0.046	0.9514	22	1,439	-0.727	0.0490
3	670	0.348	0.6842	23	2,667	0.082	0.7046
4	545	0.231	0.5322	24	4,023	0.092	0.6851
5	350	-1.147	0.0256	25	852	-0.699	0.0842
6	463	-0.608	0.1101	26	5,641	0.094	0.6947
7	2,851	-0.590	0.0291	27	680	0.602	0.2453
8	364	0.349	0.1953	28	1,135	-0.174	0.6883
9	761	-0.678	0.0638	29	1,267	0.445	0.4060
10	9,259	-0.089	0.7235	30	3,173	-0.003	0.9902
11	3,000	-0.133	0.226	31	325	0.910	0.1431
12	3,828	-0.235	0.4236	32	654	-0.632	0.1495
13	1,476	0.349	0.3829	33	9,493	0.183	0.2093
14	1,416	0.236	0.4684	34	1,009	0.810	0.4225
15	1,736	-0.211	0.5539	35	335	0.950	0.0206
16	2,111	-0.170	0.6205	36	3,849	-0.234	0.3455
17	4,765	0.043	0.8393	37	400	0.872	0.1824
18	749	-0.067	0.9079	38	9,997	-0.186	0.4076
19	2,694	0.283	0.3698	39	6,564	-0.122	0.5641
20	2,992	0.488	0.1674				

Shape parameters from the loglogistic regression in Table 10. Estimates of the loglogistic distribution's shape parameter $\ln(\gamma)$ are reported by firm. The hazard rate is monotonically decreasing when $\ln(\gamma) \geq 0$ and hump-shaped when $\ln(\gamma) < 0$. The p-values refer to the null hypothesis of $\ln(\gamma) = 0$.

Table 12: Compound effects of unemployment, income, and home ownership

Year	Median white-collar job	Median unemployed	Difference
2005	99.52	97.84	-1.68*
2006	98.99	95.56	-3.43**
2007	98.51	93.52	-4.99**
2008	98.10	91.86	-6.25**
2009	97.74	90.40	-7.34**
2010	97.41	89.09	-8.33**
2011	97.17	88.12	-9.05**

Survival probabilities estimated from the loglogistic regression in Table 10. Entries represent estimated survival probabilities in percentage points of a contract originated on 1st January 2005 in two scenarios. Scenario 1: All covariates take the sample median/mode values conditional on a main income earner working a white-collar job in middle management. Scenario 2: All covariates take the sample median/mode values conditional on an unemployed main income earner. The table shows year-end survival probabilities in the two scenarios as well as their difference. The null hypothesis of a zero-difference is tested. $^{*}p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$. See Figure 3 for an illustration.

Table 13: Robustness checks for the PRH[1]

Covariate	Level	Model 1 Hazard ratio	Model 2 Avg. marg. eff.
Contract age (quarters)			-0.006***
Nominal monthly net income of household (EUR)	(0-1499)		
	1500-1999	0.829	-0.202
	2000-2499	0.817	-0.231*
	2500-2999	0.770*	-0.268*
	3000-3999	0.651***	-0.416***
	4000+	0.469***	-0.604***
Wealth	Real estate ownership	0.816**	-0.170**
	Financial assets	0.630***	-0.407***
Occupation of main income earner	Retired	0.739*	-0.190
	In education/not in labor force	0.874	0.075
	Low blue-collar	0.750	-0.254
	High blue-collar	1.004	-0.034
	Low white-collar	1.006	-0.049
	(Middle white-collar)		
	High white-collar	0.747	-0.234*
	State-employed	1.123	0.121
	Self-employed	1.228	0.194
	Unemployed	1.860***	0.733**
Highest degree in household	No vocational training	1.682**	0.513
	Lower secondary education	1.069	0.107
	Middle secondary education	1.108	0.080
	Advanced secondary education	0.782	-0.206*
	Postsecondary education	0.888	-0.079
	(University degree)		
Year	(2005)		
	2006	0.892	-0.120
	2007	0.849	-0.153
	2008	0.766*	-0.231*
	2009	0.716**	-0.307**
	2010	0.555***	-0.439***
	2011	0.453***	-0.580***
No internet access		0.786**	-1.784**
Residence in East Germany		1.118	0.120
Age of policyholder at inception (years)	0-17	0.858	-0.133
	18-22	1.070	0.078
	23-25	1.140	0.137
	(26-30)		
	31-35	1.031	0.059
	36-40	1.012	0.024
	41-45	1.109	0.071
	46-55	1.011	-0.005
	56+	0.891	-0.110
Household structure	Single male	0.789	-0.178
	Single female	0.884	-0.097
	(Couple)		
	More than two persons	1.065	0.083
Town size	(0-20 thousand)		
	20-200 thousand	0.907	-1.014
	>200 thousand	0.877	-1.305

[1] PRH: Policy replacement hypothesis

The table shows regression results of two models after the sample has been restricted as follows. All policies are excluded whenever a purchase of another policy by the same household is observed at any time during the sampling period. Model 1: Hazard ratios from a Cox proportional hazards regression. Model 2: Average marginal effects in percentage points from a pooled panel logit. Both regressions are based on 111,518 observations; 33,108 contracts; 934 failures; 442,819 quarters at risk. Standard errors are adjusted for 16,294 household clusters. The Cox proportional hazards regression is stratified by 39 firms and in the pooled panel logit fixed effects for all 39 firms are included (not shown). $^{*}p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$. Reference levels in parentheses. Wealth is proxied by two dummy variables that indicate whether a household owns real estate and whether a household holds any financial assets other than a current or a savings account. The hazard ratio indicates the factor by which the hazard rate changes in response to a change in the covariate relative to its reference level. The average marginal effect indicates the sample average of the change in the annual lapse probability in response to such an event. For instance, an increase in monthly net income from the 0-1499 bracket to the 1500-1999 bracket will reduce the hazard rate and the annual probability of lapse c.p. by 17.1% and c.p. by 0.2 percentage points, respectively.

D. Econometric Appendix

D.1. Stratification of the Cox Proportional Hazard Rate Model

In the standard specification of the Cox proportional hazards model, the same baseline hazard rate $h_0(t)$ is assumed for all observations j. For discrete covariates taking K levels, it is possible to assume K separate baseline hazard rates, one for each level. This is called stratification. In our application, a separate baseline hazard rate is assumed for each firm $k \in \{1,\ldots,K\}$.

$$h(t_j|x_j, f_{j,k}=1) = h_k(t)exp\left(x_j'\beta\right)$$

where $f_{j,k}$ = firm dummy taking value 1 when observation j belongs to firm k,

$h_k(t)$ = firm-specific baseline hazard rate.

Note that the other covariates x_j are assumed to have the same effect across firms. Stratification is more flexible than merely including firm dummies in the vector of covariates, as the functional form of the baseline hazard rates $h_k(t)$ may vary across strata. See Kalbfleisch and Prentice (2002, p.118) for details.

D.2. Marginal Effects in the Loglogistic Model

In this section we state the marginal effect of a discrete change in one of the covariates on the survival function. In the class of accelerated failure time (AFT) models, the following data generating process is assumed:

$$\ln(t_j) = x_j'\beta + \ln(\tau_j) \tag{2}$$

where t_j = analysis time of observation j,

τ_j = random error term.

For the loglogistic specification, it is assumed that $\tau_j \sim loglogistic(\beta_0,\gamma)$. Then the analytical derivation for $S(t_j|x_j)$ is straightforward, see Cleves et al. (2010, p.273):

$$S(t_j|x_j) = \frac{1}{1+\left(t_j e^{-\beta_0 - x_j'\beta}\right)^{\frac{1}{\gamma}}} \tag{3}$$

Note that $S(t_j|x_j)$ is the probability to survive at least until time t_j if the covariates are equal to x_j throughout the entire period $(0,t_j]$. We consider a situation where the covariate vector equals $\overline{x}$ during $(0,t']$ and then switches to $\underline{x}$ for $(t',t'+4]$. We contrast this scenario with the case where the covariates are equal to $\overline{x}$ for the entire period $(0,t'+4]$. Therefore we measure marginal effects by the following expression:

$$\Delta S(t',\overline{x},\underline{x}) = S(t'|\overline{x})\frac{S(t'+4|\underline{x})}{S(t'|\underline{x})} - S(t'+4|\overline{x})$$

As we have quarterly data, four periods of analysis time equal one year of calendar time. t' denotes the point of analysis time at which the change in covariates sets in. Since all our covariates are categorical variables, we consider differences in survival probabilities rather than the partial derivative of the survival function.